

NWS Operations Proving Ground

Operational Evaluation Report

*Evaluating Digital Aviation Services and
WFO/AWC Collaboration*



NWS Operations Proving Ground
7220 NW 101st Terrace
Kansas City, MO 64153

Table of Contents

Executive Summary	1
--------------------------------	----------

FORMAL REPORT

DAS Evaluation 2016 (June).....	3
--	----------

DAS Evaluation 2016 (August).....	4
--	----------

DAS Evaluation 2017 (August).....	5
--	----------

DAS Evaluation 2018 (March)	6
--	----------

OPG Recommendations.....	7
---------------------------------	----------

Figures.....	8
---------------------	----------

Executive Summary

During the three-year period from 2016-2018, the National Weather Service (NWS) Operations Proving Ground (OPG) and Aviation Weather Testbed (AWT) collaborated on four separate experiments aimed at evaluating and improving Digital Aviation Services (DAS) in support of the NWS operational mission. DAS is the process of producing and editing aviation grids in NWS' Advanced Weather Interactive Processing System (AWIPS) Graphical Forecast Editor (GFE) framework. Local Terminal Aerodrome Forecasts (TAFs) are then created from the edited grids using the TAF Formatter within GFE. Currently all Eastern Region WFOs are using DAS operationally. In Central Region, approximately 80% of WFOs are using DAS; and all will be by the end of FY18. Less than ten offices in Southern Regions have adopted the DAS methodology, and in Western Region, only four are using DAS experimentally. The primary reason for the slower acceptance in Western Region is that the DAS process has significant shortcomings in complex terrain.

The overarching goal of the DAS evaluations was to determine the ability of the Aviation Weather Center and WFOs to effectively collaborate on the production of national-scale grids of clouds and visibility that could serve the spectrum of service needs ranging from guidance for national air space traffic management down to local TAF-driven operations. Secondary, but important, objectives included assessing the capability to edit and produce representative aviation grids from a common starting point without causing adverse workload impacts; and evaluating the representativeness of output generated from aviation grids by the AWIPS TAF Formatter. Four evaluations were conducted from 2016 through 2018.

On the OPG side, 21 NWS forecasters participated in the WFO function of the evaluations, representing WFOs in Alaskan, Western, Central, Southern, and Eastern regions. On the AWT side, almost 30 forecasters participated, representing both WFOs and CWSUs from all NWS Regions. The AWT portion of the experiments also included end-users from various divisions of the FAA, the commercial airline industry, and international weather forecasting agencies (e.g., UK Met Office, Environment Canada, and Taiwan Weather Agency).

The details described in this report are solely focused on the WFO portion of the experiments conducted in the OPG. Please contact the Aviation Weather Testbed for results associated with tools, products, and functions assessed from the national center perspective.

Collectively, critical WFO-related findings from the four DAS Evaluations were as follows:

- 1) The TAF Formatter output has improved remarkably since the first (2016) DAS evaluation. Forecasters originally noted that the formatter included far more meteorological detail than they would have included in a manual TAF, and getting the output into a format that was usable for dissemination was somewhat time-consuming. In the last (2018) evaluation forecasters felt that the TAF formatter was far more skillful in producing TAFs that are representative of what they expected to see based on the weather in their grids. Some forecasters even stated that they would like to see slightly more detail in certain circumstances. Most forecasters expressed that it is easier to delete additional unneeded information than to add missing information. However, there is a tipping point where a draft product is so lengthy that it becomes a nuisance to correct. In general, the majority of 2018 forecasters rated the representativeness of the TAFs created by the TAF formatter as "good" to "very good."

- 2) Forecasters found that NWSChat was slightly to moderately effective for meeting their needs in collaborating with the AWC to create representative ceiling and visibility grids. Forecasters unanimously agreed that the ability to share graphics in NWSChat would make collaboration much easier, both among WFOs and between WFOs and National Centers.
- 3) There was no consensus among forecasters as to whether DAS generated TAFs were more or less efficient than manually generated TAFs.
- 4) There was no consensus on whether existing DAS methodology (the use of CONSShort as a starting point) should be replaced with experimental AWC guidance.
- 5) There was no consensus on differences in quality when comparing TAFs generated from AWC guidance to TAFs generated from NBM guidance.
- 6) Forecasters felt that both NBM and AWC Guidance needed more refinement.

Based on results from the DAS evaluations, OPG recommends the following actions:

- 1) While the TAF Formatter code has demonstrated significant improvement in its ability to produce representative aviation forecasts, there are still refinements needed before its output is consistent and reliable. Whether it makes sense for these refinements to be programmed by GSD or assigned to another entity is outside the scope of this work and beyond the authority of OPG or AWT.
- 2) This series of experiments, conducted in a controlled setting with limited cases, has probably achieved its limit of practical lessons learned. To answer the remaining questions concerning the operational readiness of DAS tools and methods, the next logical step is to assess the efficacy of these procedures over a longer time frame that encompasses greater geographical diversity and a larger sample size of weather events. One possible approach would be for AWC to test real-time collaboration on national DAS grids with a collection of WFO pilot offices in a season-long, or year-long, demonstration.
- 3) NBM and AWC guidance both require continued development. A recurring theme from forecaster feedback is that blending observational information into the first few hours of output is a consistent challenge in creating accurate aviation grids. Currently, forecasters find the easiest way to achieve desired results is to blend the first guess grids with CONSShort. Forecasters will undoubtedly continue to employ that practice until the NBM proves it can produce grids comparable to, or better than, that method.

DAS Evaluation 2016 (June)

In June of 2016 the first DAS evaluation took place. Six NWS forecasters from WFOs in Western, Central, Eastern, Southern, and Alaska Regions participated in the evaluation (Figure 1). The evaluation was conducted over two weeks with three WFO forecasters participating each week on the OPG side. For this evaluation the participating forecasters did not have prior experience with DAS. Four subject matter experts (SMEs) also assisted with the evaluation.

The goals of this evaluation were to:

- Obtain forecaster feedback on the effectiveness, viability, and workload associated with generating enhanced short-term aviation grids and products necessary to support NWS aviation partners. Specific tasks included editing, saving, and publishing aviation grids modified from a common starting point; and creating TAFs derived from those grids.
- Qualitatively and statistically assess how forecaster modifications impact the AWC First Guess Cloud and Visibility Grids.
- Evaluate whether WFO modified ceiling and visibility grids can be efficiently transferred back to AWC via the LDM.

The original plan for this evaluation was to use archived weather data. This would allow OPG and AWT staff to select events that would be ideal for aviation grid editing and TAF generation. However, through extensive testing it was determined that technological challenges associated with the NDFD prevented us from developing full operational functionality. Therefore, the evaluations were conducted using real-time weather data ingest only.

Findings:

- 1) Four out of six forecasters found the grid monitor slightly useful to moderately useful in assessing the quality of forecast grids. Two forecasters did not use the Grid Monitor.
- 2) All six forecasters felt that the TAF Formatter included more meteorological detail than they would have included in a manual TAF.
- 3) There was no consensus on whether DAS-generated TAFs were more or less efficient than manual TAFs.
- 4) There was no consensus on whether DAS-generated TAF cloud layers and ceilings derived from AWC Cloud Bases were better or worse than those derived from CONSShort.
- 5) There was no consensus on whether DAS-generated TAF visibilities derived from AWC Visibilities were better or worse than those derived from CONSShort.
- 6) There was no consensus on whether DAS-generated TAFs were more efficient or less efficient to compose than manual TAFs.
- 7) There was no consensus on whether DAS-generated TAFs increased or decreased forecaster workload.

DAS Evaluation 2016 (August)

In August of 2016 the second DAS evaluation took place in the OPG. This evaluation differed from the June evaluation in both stated goals and forecaster experience levels. NWS forecasters in this evaluation were all proficient in the use of DAS procedures. The forecasters were from WFOs in Central and Eastern Regions (Figure 1), serving in offices where DAS is part of the operational routine. As with the first set of experiments, this evaluation was held over two week-long sessions, with three WFO forecasters each week, six forecasters in total.

The goals of this evaluation were to:

- Obtain forecaster feedback on the effectiveness, viability, and workload associated with generating enhanced short-term aviation grids and products necessary to support NWS aviation partners. Specific tasks included editing, saving, and publishing aviation grids modified from a common starting point; and creating TAFs derived from those grids.
- Qualitatively (OPG) and statistically (AWC) assess how forecaster modifications impact the AWC First Guess Cloud and Visibility Grids.
- Obtain feedback on the AWC-WFO collaborative process, the tools available to accomplish it, and the potential impacts on workload.
- Evaluate whether national ceiling and visibility grids can be efficiently transferred from AWC to WFOS and back, via the LDM. For the evaluation, AWC grids were used as the first guess for clouds, ceilings, and visibilities. Real-time weather was used for all exercises.

Findings:

- 1) Four out of six forecasters considered TAFs derived from standard DAS methodology (i.e., using CloudBasePrimary and percent cloud cover to determine ceilings), as superior in quality and representativeness to those generated from AWC Ceiling and Visibility guidance. Two forecasters thought the methodologies were comparable in quality.
- 2) Five out of six forecasters found that it took more time to produce DAS-generated TAFs from AWC Ceiling and Visibility guidance than from the CloudBasePrimary methodology. One forecaster thought the methodologies were comparable in the amount of time it took to produce a TAF.
- 3) Five out of six forecasters found that NWSChat was slightly to moderately effective for meeting their needs in collaborating with AWC to create representative ceiling and visibility grids. One forecaster indicated that NWSChat was neither effective nor ineffective for meeting their need for collaborating with AWC to create representative ceiling and visibility grids. All forecasters agreed that the ability to share graphics in NWSChat would make collaboration much easier.
 - a) Forecaster quote: *“The flow of information from the AWC and the OPG back to the AWC was good, but it certainly could be improved through the use of the Collaboration portion of CAVE/D2D on AWIPS2 due to the ability to share and annotate graphics. The experiment did not get to test this aspect out to its fullest extent due to some technical problems and the limitations of text based chatting.”*
- 4) There was no consensus as to whether existing DAS methodology should be replaced with experimental AWC guidance.

DAS Evaluation 2017 (August)

In August of 2017 the third DAS evaluation took place in the OPG. The evaluation was held over two weeks with three WFO forecasters participating each week in the OPG portion of the experiment – six forecasters in total. The forecasters were from WFOs in Western, Central, Eastern, and Southern Regions (Figure 1). NWS forecasters evaluated tools, methods, and the collaboration process of AWC and WFOs working together to produce a single set of ceiling and visibility grids that serve both national and local needs for aviation products and services.

The goals of this evaluation were to:

- Obtain forecaster feedback on the effectiveness, viability, and workload associated with generating enhanced short-term aviation grids and products necessary to support NWS aviation partners. Specific tasks included editing, saving, and publishing aviation grids modified from a common starting point; and creating TAFs derived from those grids. The NBM v3.0 was used as the starting point for generating 15Z TAFs; and the AWT grids were used as the starting point for generating 18Z TAFs.
- Qualitatively (OPG) and statistically (AWC) assess how forecaster modifications impact the AWC First Guess Cloud and Visibility Grids.
- Obtain feedback on the AWC-WFO collaboration process, the tools available to accomplish it, and the potential impacts on workload.
- Evaluate whether national ceiling and visibility grids can be efficiently transferred from AWC to WFOS and back, via the LDM.

Findings:

- 1) There was no consensus on differences in quality when comparing TAFs generated from AWC guidance to TAFs generated from NBM guidance.
- 2) There was no consensus on differences in the workload of producing TAFs generated from AWC guidance to producing TAFs generated from NBM guidance.
- 3) Three forecasters were extremely uncomfortable with replacing existing DAS methodology with AWC guidance. One forecaster was slightly uncomfortable with replacing existing DAS methodology with AWC guidance. Two forecasters were slightly comfortable with replacing existing DAS methodology with AWC guidance

DAS Evaluation 2018

In March of 2018, the fourth and final DAS evaluation took place in the OPG. This evaluation specifically targeted DAS for winter weather scenarios. This evaluation was one week long with three NWS forecasters from Central and Eastern Regions (Figure 1).

The goals of this evaluation was to obtain forecaster feedback on the effectiveness, viability, and workload associated with generating enhanced short-term aviation grids and products necessary to support NWS aviation partners from a common starting point, and creating TAFs derived from those grids. For some exercises, the common starting point was the NBM; for other exercises, AWC experimental guidance was used. NBM 3.1 was available for one of the days; NBM 3.0 was used for the others.) As with previous experiments, there were no historical cases in this evaluation. All exercises were conducted with forecasters editing grids in real time, based on current weather data.

Findings:

- 1) Two forecasters felt that the process of editing grids to create an automatic TAF was moderately difficult. One forecaster indicated that the process of editing grids to create an automatic TAF was moderately easy.
- 2) In general, forecasters held the opinion that representativeness of the automated TAF from the grids were good to very good.
 - a) Forecaster quote: *"I felt that the TAF Formatter does the job for the most part of taking what is put into the grids and putting it into TAF form. The biggest source of effort will be to make sure the grids tell the weather story I want it to, and project that in the TAF. Some 'post-process' editing may be necessary for individual TAF sites to make this happen."*
- 3) Forecasters agreed that automated TAFs contained slightly too little detail. This result is different from the other evaluations. In Evaluations 1-3, forecasters noted that the TAFs derived from the grids and generated by the TAF Formatter typically contained too much detail. The forecasters had to delete multiple lines of text and edit the content more often than the forecasters in Evaluation 4. Thus, the changes made to the TAF Formatter code have been overall positive.
- 4) Forecasters felt that the amount of editing that would be required after an automatic TAF was generated was slight to moderate.
- 5) Two forecasters felt that the NBM first guess guidance grids were moderately noisy. One forecaster felt that the NBM first guess guidance grids were slightly too smooth.
- 6) Two forecasters felt that the AWC first guess guidance grids were moderately noisy. One forecaster felt that the AWC first guess guidance grids were slightly too smooth.

OPG Recommendations

Over the last three years, DAS methodology and functionality has greatly improved. The number of WFOs implementing DAS methodology continues to grow. Therefore, the efficiency of the process and the representativeness of the output will be increasingly important. DAS training for the entire NWS is currently being developed. With this knowledge in mind, the results of these collaborative experiments point to three recommended actions:

- 1) While the TAF Formatter code has demonstrated significant improvement in its ability to produce representative aviation forecasts, there are still refinements needed before its output is consistent and reliable. Whether it makes sense for these refinements to be programmed by GSD or assigned to another entity is outside the scope of this work and beyond the authority of OPG or AWT.
- 2) This series of experiments, conducted in a controlled setting with limited cases, has probably achieved its limit of practical lessons learned. To answer the remaining questions concerning the operational readiness of DAS tools and methods, the next logical step is to assess the efficacy of these procedures over a longer time frame that encompasses greater geographical diversity and a larger sample size of weather events. One possible approach would be for AWC to test real-time collaboration on national DAS grids with a collection of WFO pilot offices in a season-long, or year-long, demonstration.
- 3) NBM and AWC guidance both require continued development. A recurring theme from forecaster feedback is that blending observational information into the first few hours of output is a consistent challenge in creating accurate aviation grids. Currently, forecasters find the easiest way to achieve desired results is to blend the first guess grids with CONSShort. Forecasters will undoubtedly continue to employ that practice until the NBM proves it can produce grids comparable to, or better than, that method.

Figures

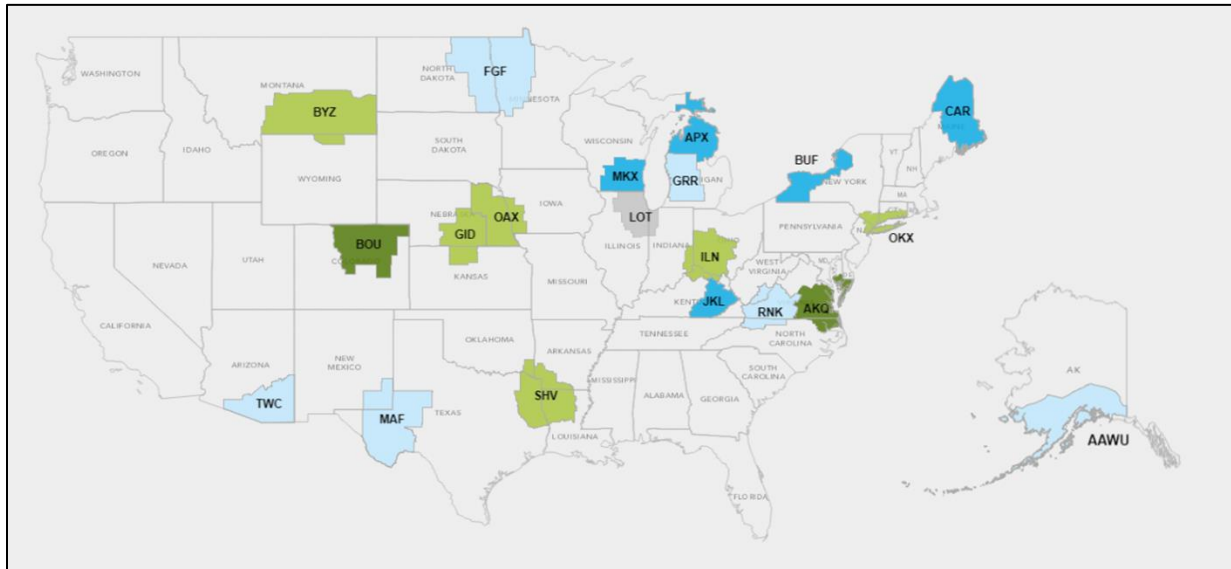


Figure 1. Colored map of forecaster's home WFO and AWU locations. The DAS evaluation for June 2016 had forecasters from AAWU, FGF, GRR, MAF, RNK, and TWC (highlighted as light blue on the map). The DAS evaluation for August 2016 had forecasters from APX, BUF, CAR, JKL, LOT, and MKX (highlighted as dark blue and gray (LOT) on the map). The DAS evaluation for 2017 had forecasters from BYZ, GID, ILN, OAX, OKX, and SHV (highlighted as light green on the map). The DAS evaluation for 2018 had forecasters from AKQ, BOU, and LOT (highlighted as dark green and gray (LOT) on the map). LOT is marked as gray because two forecasters were from that WFO.